

Characterising planktonic dinoflagellate diversity in Singapore using DNA metabarcoding

Yue Sze¹, Lilibeth N. Miranda², Tsai Min Sin^{2,†}, Danwei Huang^{1,2}

¹ Department of Biological Sciences, National University of Singapore, Singapore 117558, Singapore.

² Tropical Marine Science Institute, National University of Singapore, Singapore 119227, Singapore.

† Deceased.

Corresponding author: Danwei Huang (huangdanwei@nus.edu.sg)

Academic editor: Thorsten Stoeck | Received 19 March 2018 | Accepted 24 April 2018 | Published 17 May 2018

Abstract

Dinoflagellates are traditionally identified morphologically using microscopy, which is a time-consuming and labour-intensive process. Hence, we explored DNA metabarcoding using high-throughput sequencing as a more efficient way to study planktonic dinoflagellate diversity in Singapore's waters. From 29 minimally pre-sorted water samples collected at four locations in western Singapore, DNA was extracted, amplified and sequenced for a 313-bp fragment of the V4–V5 region in the 18S ribosomal RNA gene. Two sequencing runs generated 2,847,170 assembled paired-end reads, corresponding to 573,176 unique sequences. Sequences were clustered at 97% similarity and analysed with stringent thresholds (≥ 150 bp, ≥ 20 reads, $\geq 95\%$ match to dinoflagellates), recovering 28 dinoflagellate taxa. Dinoflagellate diversity captured includes parasitic and symbiotic groups which are difficult to identify morphologically. Richness is similar between the inner and outer West Johor Strait, but variations in community structure are apparent, likely driven by environmental differences. None of the taxa detected in a recent phytoplankton bloom along the West Johor Strait have been recovered in our samples, suggesting that background communities are distinct from bloom communities. The voluminous data obtained in this study contribute baseline information for Singapore's phytoplankton communities and prompt future research and monitoring to adopt the approach established here.

Key Words

high-throughput sequencing; Johor Strait; molecular operational taxonomic unit; phytoplankton; Singapore Strait

Introduction

Dinoflagellates (Alveolata: Dinophyceae) are a diverse and abundant group of unicellular protists found in both marine and freshwater environments. With more than 2,000 living species described (Gómez 2012), they are one of the most functionally important organisms in a variety of aquatic ecosystems (Spector 1984; Taylor et al. 2008). On top of being species rich, they have highly diversified morphologies, physiologies and biochemical properties (Hackett et al. 2004; Taylor et al. 2008). Dinoflagellates can be armoured or unarmoured (naked), based on the presence or absence of thecal plates respectively (Netzel and Dürr 1984). They play diverse ecological roles—as autotrophs, heterotrophs or mixotrophs—and can be endosymbionts of other organisms, notably the shallow-water reef-building corals, or host other endosymbionts them-

selves (Taylor et al. 2008). As part of the phytoplankton community, they are primary producers that form the base of food chains and are a large constituent of the aquatic system's food web (Spector 1984; Taylor et al. 2008).

In the marine environment, there are more than 1,500 species including both photosynthetic and non-photosynthetic types (Taylor et al. 2008, Gómez 2012). Massive proliferation and accumulation of planktonic dinoflagellates can lead to water colouration and certain species are also known to cause damage to marine ecosystems as harmful algal blooms or 'red tides' (Berdalet et al. 2016). Fish kills and other animal mortalities are common especially when the bloom involves toxin-producing dinoflagellates such as *Karenia brevis* (Landsberg et al. 2009). Incidents of paralytic shellfish poisoning by dinoflagellates such as *Alexandrium fundyense* have serious health implications higher up the food chain especially where human consumption is involved

(Tan and Lee 1986; John et al. 2015). For example, in 2008, an intense bloom in the St Lawrence Estuary was found to be responsible for the death of many fish, bird and mammal species and extremely hazardous levels of paralytic shellfish toxins produced by the dinoflagellate were detected in edible mussels collected during the event (Starr et al. 2017).

The most common and earliest method of observing phytoplankton is through light microscopy, which is time consuming, labour intensive and demands a high level of taxonomic expertise (Sinigalliano et al. 2009; Medlin 2013; McNamee et al. 2016). Some species, such as those in *Alexandrium* Halim, 1960, are relatively featureless and furthermore belong to species complexes, rendering them challenging to identify morphologically (John et al. 2005; Anderson et al. 2012; John et al. 2014). In addition, naked dinoflagellates are also difficult to identify from preserved samples. Hence, more efficient approaches to study dinoflagellates have been developed, including the use of scanning electron microscopy (Jung et al. 2010), natural fluorescence (Karlson et al. 2010), flow cytometry (Sinigalliano et al. 2009), high performance liquid chromatography (HPLC) (Gin et al. 2006) and molecular methods (Karlson et al. 2010).

Molecular techniques exploit the genetic distinction between species to identify and quantify species and have played an integral role in our understanding of the systematic positions and evolutionary relationships amongst organisms (Karlson et al. 2010). Examples of earlier molecular methods include cloning and sequencing (Zardoya et al. 1995), as well as heteroduplex mobility assay (Oldach et al. 2000). More recently, DNA barcoding has been successful in identifying dinoflagellates (Stern et al. 2010). It uses the amplified sequence from a short, standardised DNA locus as diagnostic characters for species identification (Lin et al. 2009; Shokralla et al. 2012). Loci tested for dinoflagellates include the nuclear small ribosomal subunit (18S rRNA) (Lin et al. 2006), large ribosomal subunit (28S rRNA; D1/D2 region) (Elferink et al. 2017), internal transcribed spacers (Stern et al. 2012) and the mitochondrial cytochrome c oxidase subunit I and cytochrome b (Lin et al. 2009). The mitochondrial genes cannot be easily amplified from all dinoflagellate taxa (Stern et al. 2012) and, amongst the nuclear loci, 18S rRNA remains the most well-sequenced marker in public databases due to their widespread use in dinoflagellate systematics (Mordret et al. 2018).

Improved technologies for performing high-throughput sequencing have enabled multiplexed amplification products to be sequenced from numerous samples or from an environmental sample, in a method generally known as DNA metabarcoding, which is meeting the needs of many ecologists for rapid taxon identification (Metzker 2010; Taberlet et al. 2012). DNA extracts can be amplified using primers tagged with a short barcode which labels the amplicons according to the sampling design and all amplicons can be pooled for sequencing and data traced back to individual samples (Schnell et al. 2015). DNA metabarcoding enables thousands of samples to be sequenced

simultaneously, rapidly and cost-effectively (Aylagas et al. 2016). Important ecological applications of metabarcoding include diet and biodiversity analyses, with environmental samples of various ecosystems obtained mainly from faeces (Srivathsan et al. 2015; Guilleraut et al. 2017), soil (Andersen et al. 2012; Treonis et al. 2018) or water samples (Thomsen et al. 2012; Yamamoto et al. 2017).

In the marine environment, metabarcoding has been useful for studying both benthic and planktonic eukaryotic diversity (Fonseca et al. 2010, 2014; Logares et al. 2014; de Vargas et al. 2015; Guardiola et al. 2015; Le Bescot et al. 2016; Tragin et al. 2018). Studies have targeted the 18S rRNA and successfully characterised phytoplankton communities, which are more diverse than previously thought (de Vargas et al. 2015; Le Bescot et al. 2016). Small and symbiotic species have escaped traditional microscopic detection but can now be identified via metabarcoding (Le Bescot et al. 2016, Leblanc et al. 2018). Here we use this technique to study phytoplankton communities in the coastal waters of Singapore.

The earliest studies of Singapore's marine phytoplankton communities were carried out in the mid-1900s, focusing on the Singapore Strait (Tham 1953, 1973; see Lee et al. 2015). It has long been hypothesised that the succession of phytoplankton species is likely caused by the biological history of water, stability and turbulence of the water column and monsoon-driven currents (Tham 1973). There have also been new species discoveries (Holmes 1998) and new species records for Singapore (Leong et al. 2015). More recently, studies have begun to use more advanced techniques, which include the use of extracted chlorophyll measurements (Gin et al. 2000), flow cytometry (Gin et al. 2003), HPLC pigment analysis (Gin et al. 2003, 2006), spectral fluorometric characterisation (Kuwahara and Leong 2015) and also molecular methods (Tang et al. 2007; Leong et al. 2015). The consensus of these efforts is that, amongst all phytoplankton groups, diatoms have the highest abundance except during dinoflagellate blooms (Gin et al. 2006; Leong et al. 2015). Particularly in the Johor Strait, there appears to be an inverse relationship between dinoflagellate and diatom counts (Chia et al. 1988).

There have been more than 20 positively identified species of dinoflagellates reported from Singapore's waters, but those which are known in detail have been studied mainly because they are associated with harmful algal blooms (Leong et al. 2015). These toxin-producing dinoflagellate species are commonly associated with human illnesses such as paralytic shellfish poisoning and diarrhetic shellfish poisoning (Holmes and Teo 2002). Based on phylogenetic analysis and identification using the large subunit rDNA (28S rDNA), Tang et al. (2007) presented the first evidence of ichthyotoxin production by *Alexandrium leei* and showed that *A. leei* strains in Singapore were more similar to isolates from Malaysia than to a strain from Korea. More recently, Leong et al. (2015) isolated clones of the 18S rDNA that were sequenced and compared to publicly available data from the GenBank nucleotide da-

tabase. With this method and morphological confirmation, they detected three new dinoflagellate records for Singapore—*Karenia mikimotoi*, *Karolodinium cf. australe* and *Karolodinium cf. veneficum* (Leong et al. 2015).

While more techniques are being explored, high-throughput sequencing has yet to enter the mainstream of Singapore's phytoplankton research, especially for the non-bloom baseline phytoplankton communities. Hence, the aim of this study is to take a metabarcoding approach to characterise the dinoflagellate communities at four sites along the western coast of Singapore (Fig. 1). We seek to understand the efficacy of this approach to recover dinoflagellate taxa that are typically identified microscopically and even detect species new to this locality. The data also allow a preliminary analysis of spatial and temporal variations in community composition along the West Johor Strait (WJS), to test the hypothesis that mixing between the inner and outer WJS homogenises communities along the Strait. Overall, this DNA metabarcoding study advances our understanding of dinoflagellate diversity and distribution and provides a baseline against which monitoring of phytoplankton communities in Singapore can be performed.

Material and Methods

Study sites and sampling

A total of 29 plankton samples were collected between September and December 2015 at four sites in western Singapore (Fig. 1). Along the West Johor Strait (WJS), three replicate samples per month were collected from stations at inner and outer WJS (1.45883°, 103.7202° and 1.34037°, 103.63018°, respectively; 24 samples). Along the Singapore Strait, two samples were collected from St John's Island (SJI) and three at a station off Jurong Island (JI) (1.22247°, 103.8487° and 1.29527°, 103.68202°, respectively; 5 samples).

Collections were carried out using a 15- μ m-mesh plankton net hauled vertically from 5 m depth to the water surface. From each haul, the volume of water collected was standardised using a 50-ml measuring cylinder and temporarily transferred to a plastic bottle. Each sample was filtered through an 8- μ m cellulose filter paper (Whatman, Sigma-Aldrich), which was immediately wrapped in sterilised aluminium foil, snap frozen in liquid nitrogen and stored at -30 °C prior to DNA extraction.

DNA extraction, amplification and sequencing

Each filter paper was cut into two and placed in separate 1.5-ml tubes for DNA extraction using the standard CTAB (cetyltrimethylammonium bromide) and phenol-chloroform protocol (Doyle and Doyle 1987). DNA pellets were re-suspended in 100 μ l of water, pooled between the two tubes and stored at -30°C until DNA amplification.

A 313-bp, V4-V5 region of the 18S rRNA locus was amplified using newly-designed forward primer, REL18S1F (5'-GTT GCG GTT AAA AAG CTC GTA GTT GGA-

3') and reverse primer, REL18S1R (5'-AAC AAA TCC AAG AAT TTC ACC TCT GAC-3'), which is the reverse complement of the published Dino18SF3 designed specifically for dinoflagellates (Lin et al. 2006; see also Ki 2012). This primer combination was designed based on highly conserved priming sites that could amplify a variable fragment, using a MAFFT version 7.222 (Katoh et al. 2002; Katoh and Toh 2008; Katoh and Standley 2013) alignment of known dinoflagellate species from Singapore (Leong et al. 2015). The fragment length chosen was aimed at maximising the overlap between the two paired-end reads from the Illumina MiSeq System. A unique 8-bp barcode generated using Barcode Generator 2.8 (Comai and Howell 2009) was added to the 5' end of each primer to pool samples for multiplexed sequencing (Suppl. material 1).

Three PCR replicates were carried out for each sample with a 25- μ l reaction mixture containing 2.5 μ l 10 \times reaction buffer, 2.0 μ l dNTPs, 1.0 μ l of 10 μ M forward and reverse primers, 0.2 μ l Bioready rTaq (Bulldog Bio), 2 μ l DNA diluted 5 \times or 10 \times and 16.3 μ l water. For each unique primer combination, a negative control (without DNA) was also prepared. In other words, every PCR for an actual sample was accompanied by a negative control using the same pair of barcoded primers. During the analyses, MOTUs appearing in the negative controls post-filtering would be removed from their corresponding samples. The PCR protocol comprised 1 min of initial denaturation at 94 °C, followed by 35 cycles of 45 s at 94 °C, 45 s at 53 °C and 1 min at 72 °C, ending with 3 min at 72 °C. Amplified products were quantified using the Qubit dsDNA BR Assay Kit on a Qubit 3 Fluorometer (Thermo Fisher Scientific) in order to pool approximately equal amount of PCR product for each tagged amplicon. Purification of the mixed products was performed using SureClean Plus (Bioline) following the manufacturer's protocol.

The pooled PCR products, including all negative controls, were split into two for Illumina DNA library preparation and paired-end sequencing with the Illumina MiSeq System. Approximately half of a sequencing run was targeted for each library—the first run generated read lengths of 2 \times 250 bp (MiSeq Reagent Kit v2), while the second 2 \times 300 bp (MiSeq Reagent Kit v3).

Bioinformatic pipeline

Assembly of paired-end reads was performed using Paired-End reAd mergeR (PEAR) version 0.9.6 (Zhang et al. 2014), with the criteria of 100-bp minimum overlap, minimum and maximum lengths at 150 and 330 bp respectively and Phred quality threshold of 30. Amongst all the dinoflagellate sequences from GenBank that were analysed here, few indels were found internal to the priming regions, with minimum and maximum lengths of 248 and 263 bp respectively. Therefore, our filtering criterion with respect to read length would capture all dinoflagellates.

The assembled reads were analysed in OBITools version 1.2.0 (Boyer et al. 2016). A maximum of 2-bp mismatch for primer tags was used to assign sequence

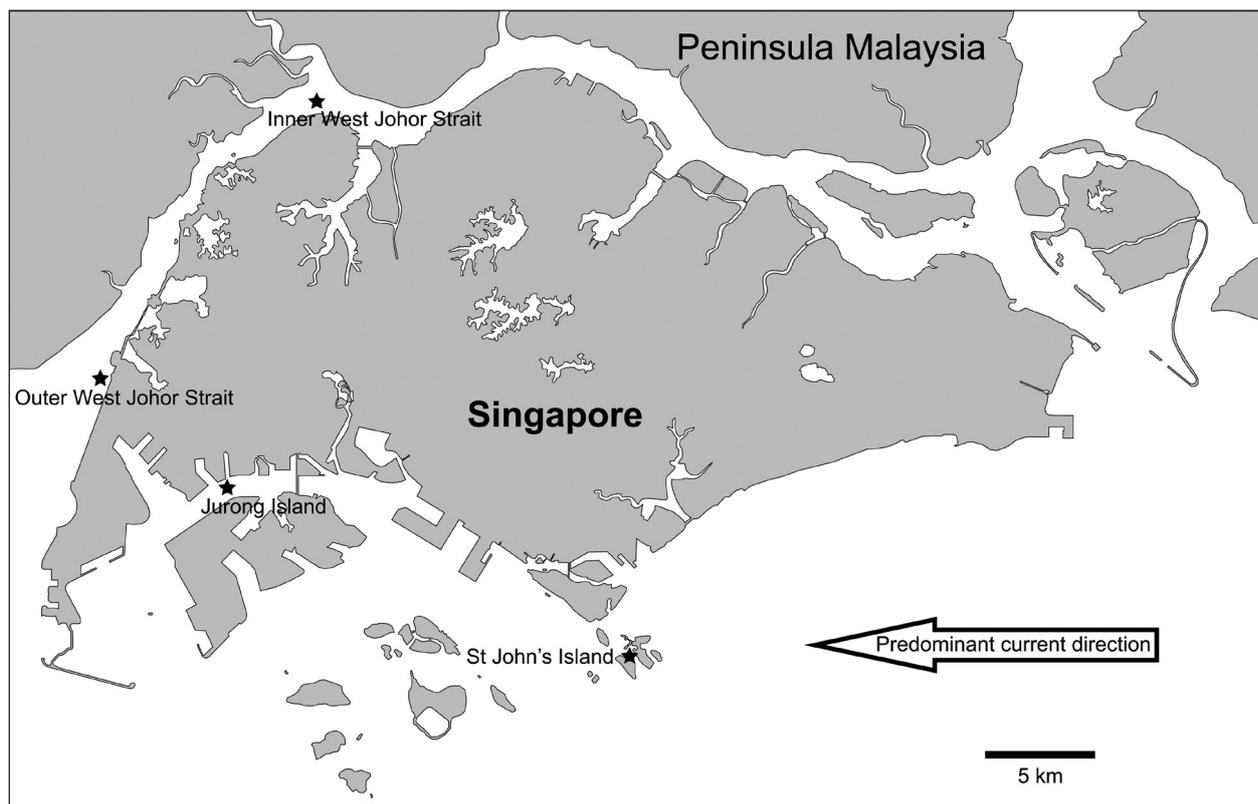


Figure 1. Sampling stations where plankton was collected for this study between September and December 2015. Arrow represents predominant current direction, stronger during the northeast monsoon (November to March) and weaker during the southwest monsoon (June–September).

records to corresponding samples with *ngsfilter*. *obianotate* followed by *obisplit* re-annotated sequence descriptions and separated them into files based on their assigned samples. Within each sample file, strictly identical sequences were grouped together and assigned a count number with *obiuniq*. *obiclean* was used to detect amplification or sequencing errors, by classifying sequence records to either ‘head’, ‘internal’ or ‘singleton’, taking into account sequence similarities and record counts. The ‘head’ is the most common sequence amongst all sequences, while a ‘singleton’ sequence is one with no other variants in the amplification product (Guardiola et al. 2015); ‘internal’ is neither of the above and most likely corresponds to a sequencing or amplification error. Only ‘head’ sequences of at least 150 bp were retained for further analysis using *obigrep* and *obisplit*.

Sequences with at least 20 reads were clustered into molecular operational taxonomic units (MOTUs) with USEARCH version 8.1.1861 (Edgar 2010) at a similarity threshold of 97% and the greedy clustering algorithm. For each cluster, the sequence with the highest number of reads was taken as the ‘centroid’, where counts of the remaining sequences with at most 3% dissimilarity were added to it. Chimeric sequences were also removed in this process. For taxonomic assignment, the representative sequence for each MOTU was searched against the GenBank sequence database using BLAST+ version 2.2.31 (Altschul et al. 1990). Only sequence matches with minimum identity of 95% to any dinoflagellate were retained for further analyses.

Data analyses

We compared dinoflagellate MOTU identities across the PCR triplicates and only retained those that met the above criteria in at least two replicates. MOTUs appearing in the negative controls were to be removed from their corresponding sample set if they met the same criteria. Retained sequences were combined first according to their sample and then site, noting their respective number of reads for each MOTU. Sequences from all samples were combined to determine the total number of unique dinoflagellate MOTUs amongst all samples.

Phylogenetic analyses were carried out to determine relationships amongst the MOTUs and previously sequenced taxa. Rhodophytes *Rhodella violacea* (Kornmann) Wehrmeyer (GenBank accession AF168624) and *Bangia atropurpurea* (Mertens ex Roth) C. Agardh (AF169339) were selected as outgroups. Taxa analysed in Leong et al. (2015) as well as published sequences from GenBank that matched $\geq 95\%$ to MOTUs obtained here were also included to construct a data matrix with 100 terminals. Sequences were aligned using MAFFT version 7.222 (Katoh et al. 2002; Katoh and Toh 2008; Katoh and Standley 2013). A neighbour-joining (NJ) tree was inferred using PAUP* version 4.0b10 (Swofford 2003), with branch supports assessed using 1,000 bootstrap pseudoreplicates (Felsenstein 1985). We also performed model-based maximum likelihood (ML) and Bayesian analyses. ML tree searches via RAxML version 8.0.9

(Stamatakis 2006, 2014; Stamatakis et al. 2008) were carried out with 50 replicates, default GTRGAMMA substitution model and 1,000 bootstrap replicates. For Bayesian analysis, we selected the most suitable substitution model using jModelTest 2.1.10 (Guindon and Gascuel 2003; Posada 2008; Darriba et al. 2012) based on the Akaike information criterion (AIC). MrBayes 3.2.6 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Ronquist et al. 2012) was used to generate four runs of 11 million Markov chain Monte Carlo iterations each, saving a tree every hundredth iteration. Convergence amongst runs was determined using Tracer 1.6 (Rambaut et al. 2014), which led to the removal of the first 50,001 trees as burn-in.

Results

The two Illumina MiSeq sequencing runs produced 2,847,170 assembled paired-end reads corresponding to 573,176 unique sequences. Further error pruning, sequence filtering and chimera removal further reduced the number of unique sequences to 268. After clustering the sequences using a 97% similarity threshold, 133 MOTUs were recovered. Filtering based on $\geq 95\%$ sequence similarity to GenBank dinoflagellate sequences resulted in 28 unique MOTUs remaining (Table 1). Sequences have been deposited into GenBank (accession numbers MH234223–MH234250). The mean number of reads per sample was 46,386 (\pm S.E. 9,743) and the mean number of MOTUs per sample was 3.79 (\pm S.E. 0.45). At the site level, considering only stations with three samples collected, there were on average 4.56 (\pm SE 1.25) dinoflagellate MOTUs per sampling station (Table 1). Our negative controls contained an average of 93 sequences per sample, but none of these met the set criteria (≥ 150 bp, ≥ 20 reads, $\geq 95\%$ match).

The phylogenetic analyses revealed deep relationships that were generally inconsistent amongst NJ, ML and Bayesian reconstructions, but these were poorly supported across all analyses. There were no topological conflicts that were supported by any of the analyses, so we discuss the well-supported nodes using the NJ tree (Fig. 2). Branch supports increased closer to the tips, thus helping to corroborate most of the BLAST matches. Of the 28 MOTUs, 14 were matched 100% to a known sequence from GenBank and most of these were strongly supported on the tree (NJ, ML bootstrap ≥ 90 , and Bayesian posterior probability = 1; e.g. OTU61 + uncultured dinoflagellate GU819712, and OTU31 + *Gonyaulax spinifera*).

OTU10, OTU37 and OTU61 showed no sequence similarity to any known species but exhibited high similarity to a limited set of ‘uncultured dinoflagellates’ (Table 1). They were phylogenetically distinct from the rest of the dinoflagellates (Fig. 2) and considered to be associated with Syndiniales, belonging to the deep-branching marine alveolates (MALV; López-García et al. 2001; Moon-van der Staay et al. 2001; Guillou et al. 2008). Other MOTUs that were not genetically similar to any known

species and were not nested within a known clade include OTU12, OTU45, OTU58, OTU96, OTU97 and OTU102. High genetic diversity was observed amongst the MOTUs nested within *Gyrodinium spirale* and *G. fusiforme* (OTU1, OTU64, OTU110, OTU123 and OTU127).

Some MOTUs could be placed in a likely taxon, such as OTU20 in *Gymnodinium* and OTU129 in *Prorocentrum*, as they had high sequence similarity with and were nested within known taxa. There were also MOTUs exhibiting exact sequence matches to specific dinoflagellate species, an indication that such species were already known. These were represented by zero branch length difference between the MOTU and the known species, such as OTU23 with *Amphidinium klebsii*, OTU73 with *Alexandrium cohorticula* and OTU81 with *Gyrodinium instriatum*.

None of the samples recovered bloom-forming dinoflagellates such as *Karodinium* and *Takayama*, which were detected in abundance during a recent phytoplankton bloom along the WJS in February 2014 (Lim et al. 2014; Leong et al. 2015). OTU98 was closely related to *Symbiodinium* sequences and was also observed during the bloom but could be part of the background community as *ex hospite* zooxanthellae.

The two most widespread MOTUs were OTU1 (identical sequence with *G. fusiforme*) and OTU20 (uncultured dinoflagellate). The former was present at both WJS sites and Jurong Island, while the latter was detected at both WJS sites and St John’s Island (Table 1). None of the MOTUs were found at all sites. OTU1 was detected at WJS during every sampling and registered consistently high read count across sites—as high as 10,000 \times the next most abundant taxon (i.e. September 2015 at inner WJS).

Nearly half of the MOTUs (13 of 28) were detected at inner and outer WJS each. There was an overlap of four MOTUs, including the abundant OTU1 (*G. fusiforme*), between the two WJS sites. Fewer MOTUs were found at Jurong Island (6 of 28) and St John’s Island (8 of 28). Two-thirds of the former’s MOTUs were unique to Jurong Island, while most of St John’s Island’s MOTUs (6 of 8) were found at WJS. Jurong Island and St John’s Island were sampled only at one time point, so results ought to be viewed with caution. At WJS, the highest MOTU richness was observed during October 2015 (Fig. 3). Particularly for outer WJS, 12 of the 13 MOTUs were detected during that month, but apart from October, the site had very low MOTU richness.

Discussion

The two Illumina MiSeq sequencing runs generated a large number of reads that likely represent the majority of dinoflagellate individuals captured on the sample filters. Overall, the use of DNA metabarcoding on minimally sorted water samples has recovered a greater diversity of dinoflagellates in Singapore than a previous study that sequenced clone isolates amplified from water samples at comparable sites (Leong et al. 2015). As the diversity estimates here are based principally on sequence analy-

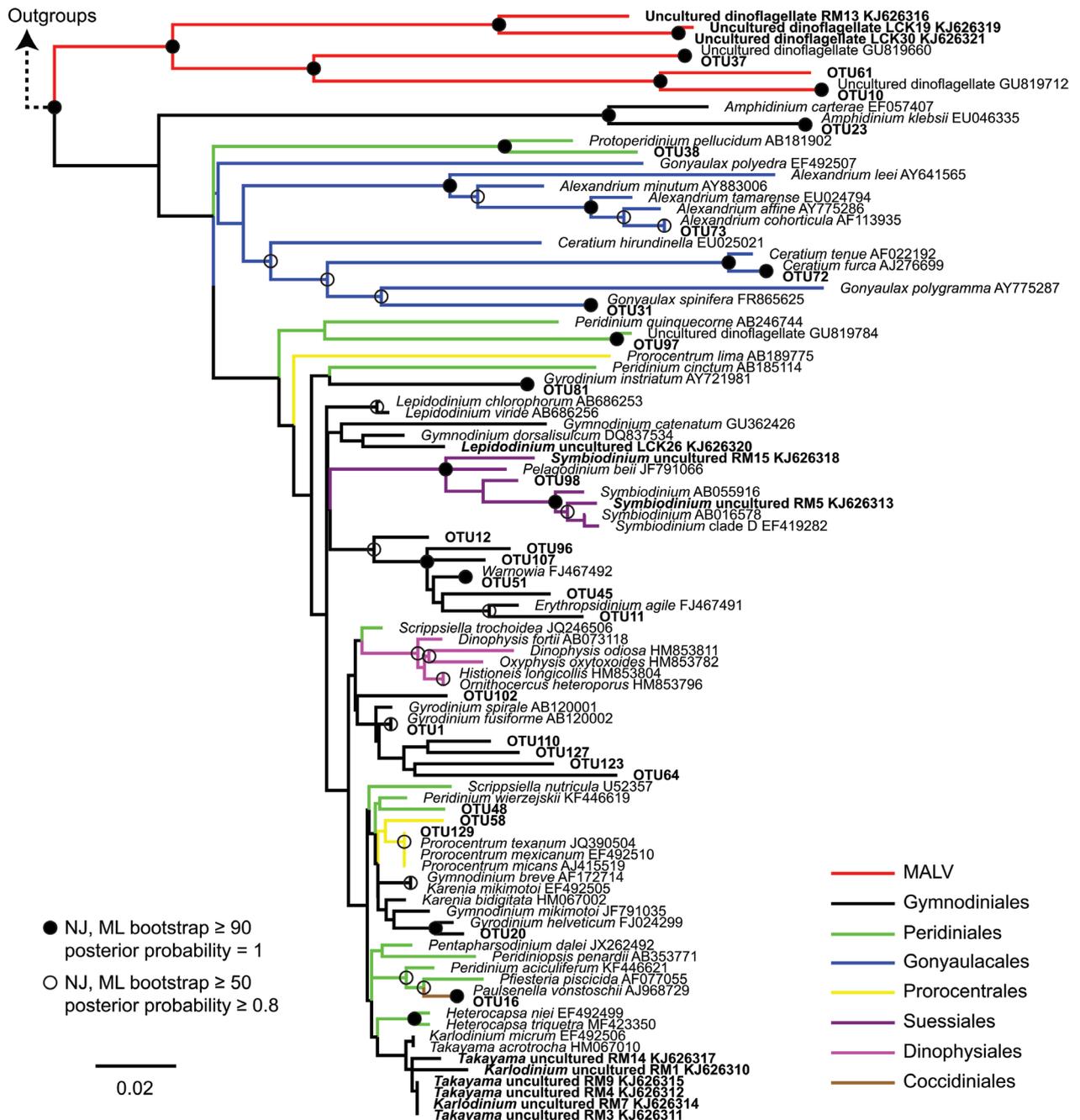


Figure 2. Neighbour-joining (NJ) tree showing 18S rRNA sequence relationships amongst dinoflagellates in Singapore (bold) and from GenBank. The 28 molecular operational taxonomic units detected in this study are denoted by prefix ‘OTU’. Dinoflagellate orders are represented by coloured branches. Bootstrap supports based on NJ and maximum likelihood (ML) methods (≥ 50), as well as Bayesian posterior probabilities (≥ 0.8) are shown as circles at the nodes.

sis, they need to be validated with other high-throughput methods such as flow cytometry and morphological examination using, for instance, scanning electron microscopy. Nevertheless, the spatial and temporal variations of community composition, shown here, form essential hypotheses for further tests of local and regional dinoflagellate distribution patterns.

After retaining sequences with $\geq 95\%$ similarity to GenBank dinoflagellate sequences, only 28 dinoflagellate MOTUs have been detected. It is worth noting that the extent of how well represented 18S rRNA dinoflagellate

sequences are in GenBank varies amongst genera. Sequences from the genus *Alexandrium* are most abundant on GenBank, with approximately 500 available. This is followed by *Prorocentrum* with 97 sequences retrieved; all other genera are represented by fewer sequences, with some having only one sequence available. Despite having the largest collection of published 18S rRNA sequences available, sequencing effort for dinoflagellates is highly skewed amongst genera. The majority of MOTUs, not assigned to dinoflagellates, indicate that the primers which have been designed for dinoflagellates also capture oth-

Table 1. Molecular operational taxonomic units of dinoflagellates recovered from four sites along the western coast of Singapore, with information about their closest GenBank matches and sequencing read counts at each site and sampling month in 2015.

| MOTU | GenBank accession number | GenBank closest match | Identity | Inner West Johor Strait | | | | Outer West Johor Strait | | | | Jurong Island | St John's Island |
|--------|--------------------------|---|----------|-------------------------|--------|--------|--------|-------------------------|------|-----|-----|---------------|------------------|
| | | | | Sep | Oct | Nov | Dec | Sep | Oct | Nov | Dec | | |
| OTU1 | MH234223 | AB120002 <i>Gyrodinium fusiforme</i> (Takano and Horiguchi 2004) | 100.0 | 367351 | 167408 | 357281 | 226843 | 147 | 4302 | 688 | 185 | 96020 | |
| OTU10 | MH234224 | GU819712 Uncultured dinoflagellate (Edgcomb et al. 2011) | 100.0 | | 1036 | 476 | 6114 | | 320 | | | 1268 | |
| OTU11 | MH234225 | AB827556 Uncultured dinoflagellate (Kok et al. 2014) | 97.7 | | | | | | 526 | | | | 1315 |
| OTU12 | MH234226 | KC488405 Uncultured dinoflagellate (Dasilva et al. 2014) | 98.5 | | 141 | | | | 4347 | | | | 99 |
| OTU16 | MH234227 | AJ968729 <i>Paulsenella vonstoschii</i> (Kühn and Medlin 2005) | 100.0 | | | | | | | | | 4435 | |
| OTU20 | MH234228 | FJ914425 Uncultured dinoflagellate | 100.0 | | 395 | | | | 4362 | 139 | | | 502 |
| OTU23 | MH234229 | EU046335 <i>Amphidinium klebsii</i> | 100.0 | | | 691 | 1968 | | | | | | |
| OTU31 | MH234230 | FR865625 <i>Gonyaulax spinifera</i> | 100.0 | | | | | | | | | 1415 | |
| OTU37 | MH234231 | GU819660 Uncultured dinoflagellate (Edgcomb et al. 2011) | 100.0 | | | | 813 | | | | | | |
| OTU38 | MH234232 | AB181899 <i>Protoperdinium pallidum</i> (Yamaguchi and Horiguchi 2005) | 95.4 | | 561 | | | | | | | | |
| OTU45 | MH234233 | EU780628 Uncultured eukaryote (Guillou et al. 2008) | 99.6 | | | | | | 257 | | | | 113 |
| OTU48 | MH234234 | FJ914426 Uncultured dinoflagellate | 98.5 | | | | 619 | | | | | | |
| OTU51 | MH234235 | FJ467492 <i>Warnowia</i> sp. (Gómez et al. 2009) | 100.0 | | | | | | | | | | 32 |
| OTU58 | MH234236 | AB694525 Uncultured dinoflagellate (Kok et al. 2012) | 99.2 | | | | | | 147 | | | | 77 |
| OTU61 | MH234237 | GU820302 Uncultured dinoflagellate (Edgcomb et al. 2011) | 100.0 | | 162 | | 23 | | | | | | |
| OTU64 | MH234238 | AB120002 <i>Gyrodinium fusiforme</i> (Takano and Horiguchi 2004) | 95.9 | | | | | | | | | 503 | |
| OTU72 | MH234239 | AJ276699 <i>Ceratium furca</i> | 99.6 | | | | | | 145 | | | | 48 |
| OTU73 | MH234240 | AF113935 <i>Alexandrium cohorticula</i> (Usup et al. 2002) | 100.0 | | | | | | 130 | | | | |
| OTU81 | MH234241 | AY721981 <i>Gyrodinium instriatum</i> | 100.0 | | | | | | 36 | | | | |
| OTU96 | MH234242 | FN598427 Uncultured dinoflagellate (Sauvadet et al. 2010) | 100.0 | | | | | | | | | | 21 |
| OTU97 | MH234243 | GU819784 Uncultured dinoflagellate (Edgcomb et al. 2011) | 99.6 | | 21 | | 21 | | | | | | |
| OTU98 | MH234244 | FJ823465 Uncultured <i>Symbiodinium</i> (Annenkova et al. 2011) | 100.0 | | | | 41 | | | | | | |
| OTU102 | MH234245 | AB827518 Uncultured dinoflagellate (Kok et al. 2014) | 97.7 | | 22 | | | | | | | | |
| OTU107 | MH234246 | FJ467492 <i>Warnowia</i> sp. (Gómez et al. 2009) | 98.1 | | | | | | 466 | | | | |
| OTU110 | MH234247 | AB120002 <i>Gyrodinium fusiforme</i> (Takano and Horiguchi 2004) | 98.8 | | | | | | 21 | | | | |
| OTU123 | MH234248 | AB120002 <i>Gyrodinium fusiforme</i> (Takano and Horiguchi 2004) | 96.5 | | | | | | | | | 380 | |
| OTU127 | MH234249 | AB120002 <i>Gyrodinium fusiforme</i> (Takano and Horiguchi 2004) | 98.0 | 34 | 62 | | | | | | | | |
| OTU129 | MH234250 | EF492510 <i>Prorocentrum mexicanum</i> (Leblond et al. 2010) EF492511 <i>Prorocentrum micans</i> (Leblond et al. 2010) JQ616822 <i>Prorocentrum rhathymum</i> (Herrera-Septúlveda et al. 2013) AJ415520 <i>Prorocentrum minimum</i> JQ390504 <i>Prorocentrum texanum</i> (Henrichs et al. 2013) | 100.0 | | | | | | 147 | | | | |

er organisms. Our database searches have indeed recovered several diatoms amongst other eukaryotes. The 18S region amplified falls within the V4–V5 region that is commonly used to broadly amplify eukaryotic 18S rDNA using universal primers (Hadziavdic et al. 2014). Despite using primer sequences that appear to be specific to dinoflagellates, the PCRs have also allowed non-specific binding. Future studies could consider adjacent priming sites to reduce non-target amplification.

As with the previous study in Singapore by Leong et al. (2015), our high-throughput sequencing recovers parasitic and symbiotic dinoflagellates which are challenging to detect using conventional microscopic sorting. The three MOTUs nested in the Syndiniales or MALV cluster (OTU10, OTU37 and OTU61) are clear outgroups of the ‘core’ dinoflagellate clade (Guillou et al. 2008; Taylor et al. 2008). While this and previous studies (e.g. Guillou et al. 2008; Leong et al. 2015) find Syndiniales to be monophyletic, as a whole, it may be paraphyletic (Strasert et al. 2018). Syndiniales are obligate parasites found throughout marine habitats from the water surface to sediments (Guillou et al. 2008) and are often highly represented in environmental samples (Kok et al. 2012). They have extremely diverse host ranges, associating with not only alveolates, but also copepods, cnidarians and even fish eggs (Chambouvet et al. 2008). Even if they are in low abundance, the community could quickly increase in population size and affect the abundance of its host, consequently impacting ecosystem functioning (Logares et al. 2015). It has only been recorded recently in Singapore’s waters during the 2014 bloom (Leong et al. 2015). Our results show that read counts of sequences closely related to Syndiniales are relatively high (Table 1)—an indication that these marine parasites are prevalent in our waters even in non-bloom conditions.

A symbiotic dinoflagellate of the genus *Symbiodinium* (OTU98) has been detected at inner WJS in December 2015. That this taxon is not more widespread and even absent in samples from the reef environment of St John’s Island is surprising because of its ubiquity and importance as endosymbionts of scleractinian corals (zooxanthellae; Tanzil et al. 2016), other marine invertebrates and protists (Takabayashi et al. 2012). They can exist as free-living cells that are detectable outside the hosts, sometimes in high densities and serve as symbiont sources to be taken up by juvenile corals or for uptake by adult corals faced with environmental stress (Littman et al. 2008; Takabayashi et al. 2012). While it is not impossible to identify them using microscopy alone, they are morphologically similar to other dinoflagellates and have often been misclassified (Littman et al. 2008). Therefore, DNA-based methods are useful for distinguishing *Symbiodinium* from other dinoflagellates. Knowing that they can be captured in our study, future research or monitoring programmes can utilise DNA metabarcoding to help track the pool of potential coral endosymbionts in Singapore waters.

By far, the most read-abundant MOTU is OTU1, which has a sequence identical to *Gyrodinium fusiforme*. While

taxon assignments based on the 18S rRNA gene may not be precise, this species is known to be widespread and has been recorded in Indonesia and the Malacca Strait (Dodge 1982; Noor et al. 2007). Present at the inner and outer WJS every month of sampling from September to December 2015, it is one of only four shared MOTUs at the two sites. Interesting, *G. fusiforme* was not detected during the 2014 *Karlodinium australe* bloom on the Singapore side of the channel (Leong et al. 2015). Furthermore, *Gyrodinium* was only one of several dinoflagellate taxa constituting a minute proportion (<0.2%) of cell densities observed on the Malaysian side (Lim et al. 2014). With the exception of the MALV and *Symbiodinium*, there appears to be mutual exclusivity between bloom and background dinoflagellate communities (Fig. 2). In particular, *Takayama* and *Karlodinium*, genera linked to mass fish mortality during the bloom (Leong et al. 2015), were not detected from our samples. There is possibility that they have been excluded from the inner WJS due to limited exchange with the Singapore Strait which connects to other regions. However, as they were also absent from the outer WJS, Jurong Island and St John’s Island, it is more likely that they were encysted and benthic during our sampling, although cysts have yet to be observed for these species (Bergholtz et al. 2005; Wang et al. 2011). Sequencing of benthic sediments from the WJS may help detect and identify cysts that can be matched to the vegetative forms in the water column (Godhe et al. 2002; Nagai et al. 2012; Gao et al. 2017).

Contrary to the hypothesis that mixing between the inner and outer WJS homogenises the communities along the Strait, considerable differences have been found between the two sampling stations (Fig. 3; Table 1). The West Johor Strait is a shallow channel bounded by the Causeway and Pulau River (Kazemi et al. 2014). Species compositional differences between the inner and outer WJS suggest that there are site-specific variations related to coastal hydrodynamics and physico-chemical properties. Surrounding water bodies and monsoonal patterns determine much of the local hydrodynamics and water parameters such as salinity, temperature and turbidity (Behera et al. 2013). The circulation of Singapore’s waters is largely driven by two dominant monsoon seasons—the northeast (November to March) and southwest (June–September) monsoons separated by two brief inter-monsoon periods (Behera et al. 2013). Our sampling period coincides only briefly with the two monsoons and the four-month sampling period is insufficient for making meaningful temporal comparisons. Nevertheless, the temporal variations appear to be related to flushing potential. Tidal mixing would reduce stratification and the hydrodynamic response to variation in freshwater inflows is dependent on the flushing time of the channel (Kazemi et al. 2014). The inner WJS is known to have low flushing potential with a high residence time of >70 days (Bayen et al. 2013) and is also affected by consistently high levels of run-off from the Johor rivers resulting in more eutrophic conditions (Sin et al. 2016). These factors could drive the slightly more uniform community structure and

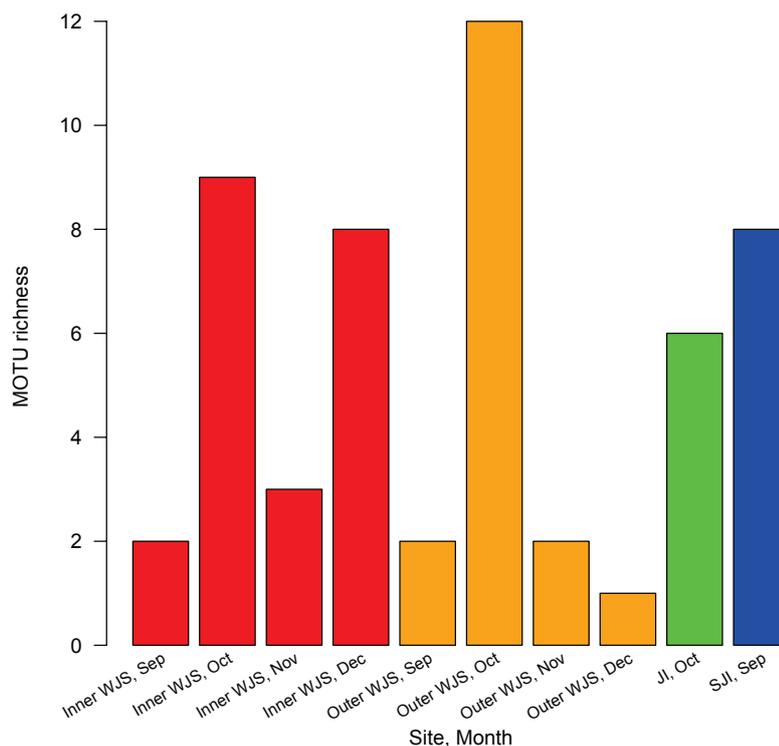


Figure 3. Number of dinoflagellate molecular operational taxonomic units (MOTUs) recovered at each of four sites, inner West Johor Strait (Inner WJS, red), outer West Johor Strait (Outer WJS, orange), Jurong Island (JI, green) and St John's Island (SJI, blue), as well as the month of sampling in 2015.

diversity over time at the inner WJS compared to the outer WJS which, being closer to the opening of the channel, would have greater tidal mixing and flushing potential.

The primary goal of all high-throughput sequencing of amplified markers is to recover accurate MOTU sequences and richness estimates from the voluminous sequence data (Nguyen et al. 2015). The number of MOTUs is dependent on the similarity threshold applied to sequence reads for delimiting taxonomic units. Sequences that are more similar than a given threshold will be grouped into the same MOTU (Blaxter 2004). The ideal similarity value used for clustering should be one that most closely approximates the diversity of a sample at the species level (Guardiola et al. 2015). At a lower similarity threshold, fewer MOTUs could be recovered, but these then become divided into separate MOTUs when a higher similarity level is used (Logares et al. 2015). The differences in the absolute number of MOTUs could be as much as six times with clustering thresholds set at 3% and 1% dissimilarities (Logares et al. 2015) and these can vary for different genetic loci (Ki 2012). Here we have attempted clustering at various dissimilarity thresholds (1–10%) to determine the sequence variabilities that show no change in the number of clusters obtained. The suitable cut-off distance ranged from 2% to 5% amongst our samples, so we used 3% as a conservative threshold so as not to overestimate diversity at each site.

The 18S rRNA is a commonly-used and well-characterised genetic marker with a highly conserved function

across all living cells. It comprises nine hypervariable regions (V1–V9) (Ki 2012; Hadziavdic et al. 2014), with V2, V4 and V9 being at sufficient interspecific variabilities for biodiversity assessments (Hadziavdic et al. 2014). Nevertheless, all eight regions of eukaryotes (lacking V6) have been targeted for sequencing (Ki 2012). Here, in order to place our MOTUs in the same phylogenetic context with known dinoflagellates from Singapore, we designed primers to target the hypervariable V4–V5 region, which is also internal to the ~600-bp marker used by Leong et al. (2015) and has a length of 313 bp for paired-end sequencing by the Illumina MiSeq System (Hadziavdic et al. 2014; Le Bescot et al. 2016; Searle et al. 2016). It is important to note that different 18S regions can lead to varying diversity estimates (Edgcomb and Stoeck 2012), so our results need to be verified in future by analysing other variable regions.

Downstream of the amplification and sequencing steps, we have attempted to account for possible errors associated with high-throughput sequencing and DNA metabarcoding. These measures include the omission of sequences <150 bp represented by fewer than 20 reads or those not matching at least 95% to published dinoflagellate sequences. Consequently, despite high read recovery, there was low read usability, with only 268 unique sequences that were eventually used for analysis. We also retained unique sequences only if they were detected in two of three PCR replicates and would potentially remove signals appearing in the negative controls. While high-throughput sequencing studies such as this are efficient in producing

large amounts of sequence data, it is important to note that there are known issues concerning Illumina-based DNA metabarcoding, including primer tag jumps, contamination and false positive detections (Esling et al. 2015). We have sought to minimise these effects through the detailed experimental design and by implementing the rigorous criteria in our bioinformatic pipeline.

Conclusion

Despite the stringent thresholds, we have recovered a higher diversity of planktonic dinoflagellates—28 MOTUs—than previously reported for Singapore waters (Leong et al. 2015). This was achieved within a short four-month sampling period with minimal pre-sorting and without isolation of specific organisms. The 15- μ m plankton net used for capturing microplankton would have omitted even smaller dinoflagellates since studies have suggested that many open water environments are dominated by small cells in the pico- (0.8–5 μ m) and nano-plankton (5–20 μ m) ranges (Gin et al. 2000; de Vargas et al. 2015). Therefore, our estimates are conservatively low. Further studies targeting greater ranges of organism size, depth and habitat are likely to detect more taxa, both known and unknown. The large and yet expanding DNA sequence database—to which the MOTUs obtained here contribute—will enable more precise matches for DNA metabarcodes locally and in the region.

Acknowledgements

We thank staff and crew of the Ecological Monitoring, Informatics and Dynamics Research Group (Tropical Marine Science Institute) and R/V *Galaxea* (St John's Island National Marine Lab), for assisting with field research; members of the Reef Ecology Laboratory for support and assistance, especially Ywee Chieh Tay for advice on data analyses; and Céline Mercier, for providing assistance regarding usage of OBITools. This study was funded by the National Research Foundation, Prime Minister's Office, Singapore under its Marine Science R&D Programme (MSRDP-P03).

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anderson DM, Alpermann TJ, Cembella AD, Collos Y, Masseret E, Montresor M (2012) The globally distributed genus *Alexandrium*: multifaceted roles in marine ecosystems and impacts on human health. *Harmful Algae* 14: 10–35. <https://doi.org/10.1016/j.hal.2011.10.012>
- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, Orlando L, Gilbert MT, Willerslev E (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21: 1996–1979. <https://doi.org/10.1111/j.1365-294X.2011.05261.x>
- Annenkova NV, Lavrov DV, Belikov SI (2011) Dinoflagellates associated with freshwater sponges from the ancient Lake Baikal. *Protist* 162: 222–236. <https://doi.org/10.1016/j.protis.2010.07.002>
- Aylagas E, Borja A, Irigoien X, Rodríguez-Ezpeleta N (2016) Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science* 3: 96. <https://doi.org/10.3389/fmars.2016.00096>
- Bayen S, Zhang H, Desai MM, Ooi SK, Kelly BC (2013) Occurrence and distribution of pharmaceutically active and endocrine disrupting compounds in Singapore's marine environment: influence of hydrodynamics and physical-chemical properties. *Environmental Pollution* 182: 1–8. <https://doi.org/10.1016/j.envpol.2013.06.028>
- Behera MR, Chun C, Palani S, Tkalic P (2013) Temporal variability and climatology of hydrodynamic, water property and water quality parameters in the West Johor Strait of Singapore. *Marine Pollution Bulletin* 77: 380–395. <https://doi.org/10.1016/j.marpolbul.2013.09.043>
- Berdalet E, Fleming LE, Gowen R, Davidson K, Hess P, Backer LC, Moore SK, Hoagland P, Enevoldsen H (2016) Marine harmful algal blooms, human health and wellbeing: challenges and opportunities in the 21st century. *Journal of the Marine Biological Association of the United Kingdom* 96: 61–91. <https://doi.org/10.1017/S0025315415001733>
- Bergholtz T, Daugbjerg N, Moestrup Ø, Fernandez-Tejedor M (2005) On the identity of *Karodinium veneficum* and description of *Karodinium armiger* sp. nov. (Dinophyceae), based on light and electron microscopy, nuclear-encoded LSU rDNA, and pigment composition. *Journal of Phycology*, 42: 170–193. <https://doi.org/10.1111/j.1529-8817.2006.00172.x>
- Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B-Biological Sciences* 359: 669–679. <https://doi.org/10.1098/rstb.2003.1447>
- Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016) OBITools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* 16: 176–182. <https://doi.org/10.1111/1755-0998.12428>
- Chambouvet A, Morin P, Marie D, Guillou L (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* 322: 1254–1257. <https://doi.org/10.1126/science.1164387>
- Chia LS, Khan H, Chou LM (1988) The coastal environment profile of Singapore. ICLARM, Manila, 1–91.
- Comai L, Howell T (2012) Barcode Generator. Version 2.8. Available from: http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator
- Cuvelier ML, Allen AE, Monier A, McCrow JP, Messie M, Tringe SG, Woyke T, Welsh RM, Ishoey T, Lee J-H, Binder BJ, DuPont CL, Latasa M, Guigand C, Buck KR, Hilton J, Thiagarajan M, Caler E, Read B, Lasken RS, Chavez FP, Worden AZ (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences USA* 107: 14679–14684. <https://doi.org/10.1073/pnas.1001665107>
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772. <https://doi.org/10.1038/nmeth.2109>
- Dasilva CR, Li WKW, Lovejoy C (2014) Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean. *Journal of Plankton Research* 36: 344–363. <https://doi.org/10.1093/plankt/fbt123>

- de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Luke J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348: 1261605. <https://doi.org/10.1126/science.1261605>
- Dodge JD (1982) Marine dinoflagellates of the British Isles. Her Majesty's Stationery Office, London, 1–303.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin* 19: 11–15.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, Suarez P, Varela R, Epstein S (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME Journal* 5: 1344–1356. <https://doi.org/10.1038/ismej.2011.6>
- Edgcomb V, Stoeck T (2012) Massively parallel tag sequencing unveils the complexity of marine protistan communities in oxygen-depleted habitats. In: Harbers M, Kahl G (Eds) *Tag-based Next Generation Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 171–183. <https://doi.org/10.1002/9783527644582.ch11>
- Elferink S, Neuhaus S, Wohrlab S, Toebe K, Voß D, Gottschling M, Lundholm N, Krock B, Koch BP, Zielinski O, Cembella A, John U (2017) Molecular diversity patterns among various phytoplankton size-fractions in West Greenland in late summer. *Deep Sea Research Part I: Oceanographic Research Papers* 121: 54–69. <https://doi.org/10.1016/j.dsr.2016.11.002>
- Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research* 43: 2513–2524. <https://doi.org/10.1093/nar/gkv107>
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>
- Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lamshead PJD, Thomas WK, Creer S (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* 1: 98. <https://doi.org/10.1038/ncomms1095>
- Fonseca VG, Carvalho GR, Nichols B, Quince C, Johnson HF, Neill SP, Lamshead JD, Thomas WK, Power DM, Creer S (2014) Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography* 23: 1293–1302. <https://doi.org/10.1111/geb.12223>
- Gao Y, Fang H, Dong Y, Li H, Pu C, Zhan A (2017) An improved method for the molecular identification of single dinoflagellate cysts. *PeerJ* 5: e3224. <https://doi.org/10.7717/peerj.3224>
- Guillerault N, Bouletreau S, Iribar A, Valentini A, Santoul F (2017) Application of DNA metabarcoding on faeces to identify European catfish *Silurus glanis* diet. *Journal of Fish Biology* 90: 2214–2219. <https://doi.org/10.1111/jfb.13294>
- Gin KYH, Lin X, Zhang S (2000) Dynamics and size structure of phytoplankton in the coastal waters of Singapore. *Journal of Plankton Research* 22: 1465–1484. <https://doi.org/10.1093/plankt/22.8.1465>
- Gin KYH, Zhang S, Lee YK (2003) Phytoplankton community structure in Singapore's coastal waters using HPLC pigment analysis and flow cytometry. *Journal of Plankton Research* 25: 1507–1519. <https://doi.org/10.1093/plankt/fbg112>
- Gin KYH, Holmes MJ, Zhang S, Lin X (2006) Phytoplankton structure in the tropical port waters of Singapore. In: Wolanski E (Ed.) *The environment of Asia Pacific harbors*. Springer, The Netherlands, 347–375. https://doi.org/10.1007/1-4020-3655-8_21
- Godhe A, Rehnstam-Holm AS, Karunasagar I, Karunasagar I (2002) PCR detection of dinoflagellate cysts in field sediment samples from tropic and temperate environments. *Harmful Algae* 1: 361–373. [https://doi.org/10.1016/S1568-9883\(02\)00053-7](https://doi.org/10.1016/S1568-9883(02)00053-7)
- Gómez F (2012) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Océánides* 27: 65–140.
- Gómez F, López-García P, Moreira D (2009) Molecular phylogeny of the ocelloid-bearing dinoflagellates *Erythrospidinium* and *Warnowia* (Warnowiaceae, Dinophyceae). *Journal of Eukaryotic Microbiology* 56: 440–445. <https://doi.org/10.1111/j.1550-7408.2009.00420.x>
- Guardiola M, Uriz MJ, Taberlet P, Coissac E, Wangenstein OS, Turon X (2015) Deep-sea, deep-sequencing: metabarcoding extracellular DNA from sediments of marine canyons. *PLoS ONE* 10: e0139633. <https://doi.org/10.1371/journal.pone.0139633>
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704. <https://doi.org/10.1080/10635150390235520>
- Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology* 10: 3349–3365. <https://doi.org/10.1111/j.1462-2920.2008.01731.x>
- Hackett JD, Anderson DM, Erdner DL, Bhattacharya D (2004) Dinoflagellates: a remarkable evolutionary experiment. *American Journal of Botany* 91: 1523–1534. <https://doi.org/10.3732/ajb.91.10.1523>
- Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* 9: e87624. <https://doi.org/10.1371/journal.pone.0087624>
- Henrichs DW, Scott PS, Steidinger KA, Errera RM, Abraham A, Campbell L (2013) Morphology and phylogeny of *Prorocentrum texanum* sp. nov. (Dinophyceae): a new toxic dinoflagellate from the Gulf of Mexico coastal waters exhibiting two distinct morphologies. *Journal of Phycology* 49: 143–155. <https://doi.org/10.1111/jpy.12030>
- Herrera-Sepúlveda A, Hernandez-Saavedra NY, Medlin LK, West N (2013) Capillary electrophoresis finger print technique (CE-SSCP): an alternative tool for the monitoring activities of HAB species in Baja California Sur Coastal. *Environmental Science and Pollution Research* 20: 6863–6871. <https://doi.org/10.1007/s11356-012-1033-7>
- Holmes MJ (1998) *Gambierdiscus yasumotoi* sp. nov. (Dinophyceae), a toxic benthic dinoflagellate from southeastern Asia. *Journal of Phycology* 34: 661–668. <https://doi.org/10.1046/j.1529-8817.1998.340661.x>
- Holmes MJ, Teo SLM (2002) Toxic marine dinoflagellates in Singapore waters that cause seafood poisonings. *Clinical and Experimental Pharmacology and Physiology* 29: 829–836. <https://doi.org/10.1046/j.1440-1681.2002.03724.x>

- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- John U, Medlin LK, Groben R (2005) Development of specific rRNA probes to distinguish between geographic clades of the *Alexandrium tamarense* species complex. *Journal of Plankton Research* 27: 199–204. <https://doi.org/10.1093/plankt/fbh160>
- John U, Litaker RW, Montresor M, Murray S, Brosnahan ML, Anderson DM (2014) Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: the introduction of five species with emphasis on molecular-based (rDNA) classification. *Protist* 165: 779–804. <https://doi.org/10.1016/j.protis.2014.10.001>
- John U, Tillmann U, Hülskötter J, Alpermann TJ, Wohlrab S, Van de Waal DB (2015) Intraspecific facilitation by allelochemical mediated grazing protection within a toxigenic dinoflagellate population. *Proceedings of the Royal Society B—Biological Sciences* 282: 20141268. <https://doi.org/10.1098/rspb.2014.1268>
- Jung SW, Joo HM, Park JS, Lee JH (2010) Development of a rapid and effective method for preparing delicate dinoflagellates for scanning electron microscopy. *Journal of Applied Phycology* 22: 313–317.
- Karlson B, Godhe A, Cusack C, Bresnan E (2010) Introduction to methods for quantitative phytoplankton analysis. In: Karlson B, Cusack C, Bresnan E (eds) *Microscopic and molecular methods for quantitative phytoplankton analysis*. UNESCO, Paris, 5–12. <https://doi.org/10.1007/s10811-009-9461-6>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780. <https://doi.org/10.1093/molbev/mst010>
- Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298. <https://doi.org/10.1093/bib/bbn013>
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kazemi Z, Hashim NB, Aslani H, Liu Z, Craig PM, Chung DH, Ismail M (2014) Calibration of hydrodynamic modeling in western part of Johor Strait, Malaysia. *International Journal of Environmental Research* 8: 591–600.
- Ki JS (2012) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *Journal of Applied Phycology* 24: 1035–1043. <https://doi.org/10.1007/s10811-011-9730-z>
- Kok SP, Kikuchi T, Toda T, Kurosawa N (2012) Diversity analysis of protistan microplankton in Sagami Bay by 18S rRNA gene clone analysis using newly designed PCR primers. *Journal of Oceanography*, 68: 599–613. <https://doi.org/10.1007/s10872-012-0121-0>
- Kok SP, Kikuchi T, Toda T, Kurosawa N (2012) Diversity and community dynamics of protistan microplankton in Sagami Bay revealed by 18S rRNA gene clone analysis. *Plankton and Benthos Research* 7: 75–86. <https://doi.org/10.3800/pbr.7.75>
- Kok SP, Kikuchi T, Toda T, Kurosawa N (2014) The protistan microplankton community along the Kuroshio Current revealed by 18S rRNA gene clone analysis: a case study of the differences in distribution interplay with ecological variability. *Plankton and Benthos Research* 9: 71–82. <https://doi.org/10.3800/pbr.9.71>
- Kühn SF, Medlin LK (2005) The systematic position of the parasitoid marine dinoflagellate *Paulsenella vonstoschii* (Dinophyceae) inferred from nuclear-encoded small subunit ribosomal DNA. *Protist* 156: 393–398. <https://doi.org/10.1016/j.protis.2005.09.002>
- Kuwahara VS, Leong SCY (2015) Spectral fluorometric characterization of phytoplankton types in the tropical coastal waters of Singapore. *Journal of Experimental Marine Biology and Ecology* 466: 1–8. <https://doi.org/10.1016/j.jembe.2015.01.015>
- Landsberg JH, Flewelling LJ, Naar J (2009) *Karenia brevis* red tides, brevetoxins in the food web, and impacts on natural resources: decadal advancements. *Harmful Algae* 8: 598–607. <https://doi.org/10.1016/j.hal.2008.11.010>
- Le Bescot N, Mahé F, Audic S, Dimier C, Garet M-J, Poulain J, Wincker P, de Vargas C, Siano R (2016) Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environmental Microbiology* 18: 609–626. <https://doi.org/10.1111/1462-2920.13039>
- Leblanc K, Quéguiner B, Diaz F, Cornet V, Michel-Rodriguez M, de Madron XD, Bowler C, Malviya S, Thyssen M, Grégori G, Rembauville M, Grosso O, Poulain J, de Vargas C, Pujo-Pay M, Conan P (2018) Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nature Communications* 9: 953. <https://doi.org/10.1038/s41467-018-03376-9>
- Leblond JD, Lasiter AD, Li C, Logares R, Rengefors K, Evens TJ (2010) A data mining approach to dinoflagellate clustering according to sterol composition: correlations with evolutionary history. *International Journal of Data Mining and Bioinformatics* 4: 431–451. <https://doi.org/10.1504/IJDMB.2010.034198>
- Lee AC, Baula IU, Miranda LN, Sin TM (2015) A photographic guide to the marine algae of Singapore. Tropical Marine Science Institute, National University of Singapore, Singapore, 1–201.
- Leong SCY, Lim LP, Chew SM, Kok JWK, Teo SLM (2015) Three new records of dinoflagellates in Singapore’s coastal waters, with observations on environmental conditions associated with microalgal growth in the Johor Straits. *Raffles Bulletin of Zoology* S31: 24–36.
- Lim HC, Leaw CP, Tan TH, Kon NF, Yek LH, Hii KS, Teng ST, Razali RM, Usup G, Iwataki M, Lim PT (2014) A bloom of *Karlodinium australe* (Gymnodiniales, Dinophyceae) associated with mass mortality of cage-cultured fishes in West Johor Strait, Malaysia. *Harmful Algae* 40: 51–62. <https://doi.org/10.1016/j.hal.2014.10.005>
- Lin S, Zhang H, Hou Y, Miranda L, Bhattacharya D (2006) Development of a dinoflagellate-oriented PCR primer set leads to detection of picoplanktonic dinoflagellates from Long Island Sound. *Applied Environmental Microbiology* 72: 5626–5630. <https://doi.org/10.1128/AEM.00586-06>
- Lin S, Zhang H, Hou Y, Zhuang Y, Miranda L (2009) High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial *cox1* and *cob* genes for dinoflagellate DNA barcoding. *Applied Environmental Microbiology* 75: 1279–1290. <https://doi.org/10.1128/AEM.01578-08>
- Littman RA, van Oppen MJH, Willis BL (2008) Methods for sampling free-living *Symbiodinium* (zooxanthellae) and their distribution and abundance at Lizard Island (Great Barrier Reef). *Journal of Experimental Marine Biology and Ecology* 364: 48–53. <https://doi.org/10.1016/j.jembe.2008.06.034>
- Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Gobet A, Kooistra WH, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Stoeck T, Santini S, Siano R, Wincker P, Zingone A, Richards TA, de Vargas C, Massana R (2014)

- Patterns of rare and abundant marine microbial eukaryotes. *Current Biology* 24: 813–821. <https://doi.org/10.1016/j.cub.2014.02.050>
- Logares R, Mangot JF, Massana R (2015) Rarity in aquatic microbes: placing protists on the map. *Research in Microbiology* 166: 831–841. <https://doi.org/10.1016/j.resmic.2015.09.009>
- López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409: 603–607. <https://doi.org/10.1038/35054537>
- McNamee SE, Medlin LK, Kegel J, McCoy GR, Raine R, Barra L, Ruggiero MV, Kooistra WHCF, Montresor M, Hagstrom J, Blanco EP, Graneli E, Rodríguez F, Escalera L, Reguera B, Dittami S, Edvardsen B, Taylor J, Lewis JM, Pazos Y, Elliott CT, Campbell K (2016) Distribution, occurrence and biotoxin composition of the main shellfish toxin producing microalgae within European waters: A comparison of methods of analysis. *Harmful Algae* 55: 112–120. <https://doi.org/10.1016/j.hal.2016.02.008>
- Medlin L (2013) Molecular tools for monitoring harmful algal blooms. *Environmental Science and Pollution Research* 20: 6683–6685. <https://doi.org/10.1007/s11356-012-1195-3>
- Metzker ML (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics* 11: 31–46. <https://doi.org/10.1038/nrg2626>
- Moon-van der Staay SY, Watchter RD, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409: 607–610. <https://doi.org/10.1038/35054541>
- Mordret S, Piredda R, Vaulot D, Montresor M, Kooistra WHCF, Sarno D (2018) DinoREF: a curated dinoflagellate (Dinophyceae) reference database for the 18S rRNA gene. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.12781>
- Nagai S, Yamamoto K, Hata N, Itakura S (2012) Study of DNA extraction methods for use in loop-mediated isothermal amplification detection of single resting cysts in the toxic dinoflagellates *Alexandrium tamarense* and *A. catenella*. *Marine Genomics* 7: 51–56. <https://doi.org/10.1016/j.margen.2012.03.002>
- Netzel H, Dürr G (1984) Dinoflagellate cell cortex. In: Spector DL (Ed.) *Dinoflagellates*. Academic Press, San Diego, 43–105. <https://doi.org/10.1016/B978-0-12-656520-1.50007-9>
- Nguyen NH, Smith D, Peay K, Kennedy P (2015) Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist* 205: 1389–1393. <https://doi.org/10.1111/nph.12923>
- Noor NM, Anton A, Amin NM (2007) Biodiversity of dinoflagellates in the coastal waters off Malacca, Peninsular Malaysia. *Borneo Science* 21: 12–18.
- Oldach DW, Delwiche CF, Jakobsen KS, Tengs T, Brown EG, Kempton JW, Schaefer EF, Bowers HA, Glasgow HB Jr, Burkholder JM, Steidinger KA, Rublee PA (2000) Heteroduplex mobility assay-guided sequence discovery: elucidation of the small subunit (18S) rDNA sequences of *Pfisteria piscicida* and related dinoflagellates from complex algal culture and environmental sample DNA pools. *Proceedings of the National Academy of Sciences USA* 97: 4303–4308. <https://doi.org/10.1073/pnas.97.8.4303>
- Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256. <https://doi.org/10.1093/molbev/msn083>
- Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer: MCMC Trace Analysis Tool. Version 1.6. Available from: <http://beast.bio.ed.ac.uk/Tracer>
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Sauvadet AL, Gobet A, Guillou L (2010) Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific deep hydrothermal habitats. *Environmental Microbiology* 12: 2946–2964. <https://doi.org/10.1111/j.1462-2920.2010.02272.x>
- Schnell IB, Bohmann K, Gilbert MTP (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* 15: 1289–1303. <https://doi.org/10.1111/1755-0998.12402>
- Searle D, Sible E, Cooper A, Putonti C (2016) 18S rDNA dataset profiling microeukaryotic populations within Chicago area nearshore waters. *Data in Brief* 6: 526–529. <https://doi.org/10.1016/j.dib.2015.12.042>
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21: 1794–1805. <https://doi.org/10.1111/j.1365-294X.2012.05538.x>
- Sin TM, Ang HP, Buurman J, Lee AC, Leong YL, Ooi SK, Steinberg P, Teo SLM (2016) The urban marine environment of Singapore. *Regional Studies in Marine Science* 8: 331–339. <https://doi.org/10.1016/j.rsma.2016.01.011>
- Sinigalliano CD, Winshell J, Guerrero MA, Scorzetti G, Fell JW, Eaton RW, Brand L, Rein KS (2009) Viable cell sorting of dinoflagellates by multiparametric flow cytometry. *Phycologia* 48: 249–257. <https://doi.org/10.2216/08-51.1>
- Spector DL (1984) Dinoflagellates: an introduction. In: Spector DL (Ed.) *Dinoflagellates*. Academic Press, San Diego, 1–15. <https://doi.org/10.1016/B978-0-12-656520-1.50005-5>
- Srivathsan A, Sha JCM, Vogler AP, Meier R (2015) Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources* 15: 250–261. <https://doi.org/10.1111/1755-0998.12302>
- Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771. <https://doi.org/10.1080/10635150802429642>
- Starr M, Lair S, Michaud S, Scarratt M, Quilliam M, Lefavre D, Robert M, Wotherspoon A, Michaud R, Ménard N, Sauvé G, Lessard S, Béland P, Measures L (2017) Multispecies mass mortality of marine fauna linked to a toxic dinoflagellate bloom. *PLoS ONE* 12: e0176299. <https://doi.org/10.1371/journal.pone.0176299>
- Stern RF, Horak A, Andrew RL, Coffroth MA, Andersen RA, Küpper FC, Jameson I, Hoppenrath M, Véron B, Kasai F, Brand J, James ER, Keeling PJ (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. *PLoS ONE* 5: e13991. <https://doi.org/10.1371/journal.pone.0013991>
- Stern RF, Andersen RA, Jameson I, Küpper FC, Coffroth MA, Vaulot D, Le Gall F, Véron B, Brand JJ, Skelton H, Kasai F, Lilly EL, Keeling PJ (2012) Evaluating the ribosomal internal transcribed spacer (ITS)

- as a candidate dinoflagellate barcode marker. *PLoS ONE* 7: e42780. <https://doi.org/10.1371/journal.pone.0042780>
- Strassert JFH, Karnkowska A, Hehenberger E, del Campo J, Kolisko M, Okamoto N, Burki F, Janoušková J, Poirier C, Leonard G, Hallam SJ, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2018) Single cell genomics of uncultured marine alveolates shows paraphyly of basal dinoflagellates. *ISME Journal* 12: 304–308. <https://doi.org/10.1038/ismej.2017.167>
- Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. <http://paup.csit.fsu.edu>
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21: 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Takabayashi M, Adams LM, Pochon X, Gates RD (2012) Genetic diversity of free-living *Symbiodinium* in surface water and sediment of Hawai'i and Florida. *Coral Reefs* 31: 157–167. <https://doi.org/10.1007/s00338-011-0832-5>
- Takano Y, Horiguchi T (2004) Surface ultrastructure and molecular phylogenetics of four unarmored heterotrophic dinoflagellates, including the type species of the genus *Gyrodinium* (Dinophyceae). *Phycological Research* 52: 107–116. <https://doi.org/10.1111/j.1440-1835.2004.tb00319.x>
- Tan CT, Lee EJ (1986) Paralytic shellfish poisoning in Singapore. *Annals of the Academy of Medicine, Singapore* 15: 77–79.
- Tang YZ, Kong L, Holmes MJ (2007) Dinoflagellate *Alexandrium leei* (Dinophyceae) from Singapore coastal waters produces a water-soluble ichthyotoxin. *Marine Biology* 150: 541–549. <https://doi.org/10.1007/s00227-006-0396-z>
- Tanzil JTI, Ng PKA, Tey YQ, Tan HYB, Yun YE, Huang D (2016) A preliminary characterisation of *Symbiodinium* diversity in some common corals from Singapore. *COSMOS* 12: 15–27. <https://doi.org/10.1142/S0219607716500014>
- Taylor FJR, Hoppenrath M, Saldarriaga JF (2008) Dinoflagellate diversity and distribution. *Biodiversity Conservation* 17: 407–418. <https://doi.org/10.1007/s10531-007-9258-3>
- Tham AK (1953) A preliminary study of the physical, chemical and biological characteristics of Singapore Straits. HM Stationery Off (London): 1–65.
- Tham AK (1973) Seasonal distribution of the plankton in Singapore straits. Special Publication dedicated to NK Panikkar, Marine Biological Association of India 60: 60–73.
- Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE* 7: e41732. <https://doi.org/10.1371/journal.pone.0041732>
- Tragin M, Zingone A, Vault D (2018) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environmental Microbiology* 20: 506–520. <https://doi.org/10.1111/1462-2920.13952>
- Treonis AM, Unangst SK, Kepler RM, Buyer JS, Cavigelli MA, Mirsky SB, Maul JE (2018) Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. *Scientific Reports* 8: 2004. <https://doi.org/10.1038/s41598-018-20366-5>
- Usup G, Leaw CP, Ahmad A, Lim PT (2002) Phylogenetic relationship of *Alexandrium tamiyavanichii* (Dinophyceae) to other *Alexandrium* species based on ribosomal RNA gene sequences. *Harmful Algae* 1: 59–68. [https://doi.org/10.1016/S1568-9883\(02\)00003-3](https://doi.org/10.1016/S1568-9883(02)00003-3)
- Wang H, Lu D, Huang H, Göbel J, Dai X, Xia P (2011) First observation of *Karlodinium veneficum* from the East China Sea and the coastal waters of Germany. *Acta Oceanologica Sinica* 30: 112–121. <https://doi.org/10.1007/s13131-011-0168-6>
- Yamaguchi A, Horiguchi T (2005) Molecular phylogenetic study of the heterotrophic dinoflagellate genus *Protoberidinium* (Dinophyceae) inferred from small subunit rRNA gene sequences. *Phycological Research* 53: 30–42. <https://doi.org/10.1111/j.1440-1835.2005.tb00355.x>
- Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M (2018) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports* 7: 40368. <https://doi.org/10.1038/srep40368>
- Zardoya R, Costas E, López-Rodas V, Garrido-Pertierra A, Bautista JM (1995) Revised dinoflagellate phylogeny inferred from molecular analysis of large-subunit ribosomal RNA gene sequences. *Journal of Molecular Evolution* 41: 637–645. <https://doi.org/10.1007/BF00175822>
- Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620. <https://doi.org/10.1093/bioinformatics/btt593>

Supplementary material 1

8-bp barcode and primer sequences used for multiplexed sequencing

Authors: Yue Sze, Lilibeth N. Miranda, Tsai Min Sin, Danwei Huang

Data type: Sequencing data

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODBL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/mbmg.2.245136.suppl1>